

# DOCUMENT RESUME

ED 253 564

TM 850 104

**AUTHOR** Egelston, Richard L.; Heinegg, Rosemarie  
**TITLE** Longitudinal Variability of Status Scores Among Students with Varying Levels of Achievement.  
**PUB DATE** Oct 84  
**NOTE** 6p.; Paper presented at the Annual Meeting of the Northeastern Educational Research Association (Ellenville, NY, October 1984).  
**PUB TYPE** Speeches/Conference Papers (150) -- Reports - Research/Technical (143)  
**EDRS PRICE** MF01/PC01 Plus Postage.  
**DESCRIPTORS** Academic Achievement; \*Achievement Tests; Basic Skills; Cohort Analysis; Correlation; Elementary Education; Longitudinal Studies; Norm Referenced Tests; \*Standardized Tests; \*Test Reliability; \*True Scores  
**IDENTIFIERS** Iowa Tests of Basic Skills

## ABSTRACT

The stability of annual normal curve equivalent (NCE) on the Iowa Test of Basic Skills over a period of six years was established in a previous study of status (position within a reference group). This study investigated the relationship between status stability and student achievement level. Third grade students in a suburban school district who continued through grade 8 in the same class, formed three cohort groups (n=153, N=148, n=148) for this investigation. Raw scores from six consecutive years were converted to percentile ranks, using national norms, and were then transformed to normal curve equivalents for the Reading Comprehension subtest and the Total Mathematics score. Correlations between means and variancer suggested a non-linear relation. When achievement levels were blocked into deciles, the within-decile variation was very large when compared to the means, and obscured any relation between achievement level and within-subject variation, although a tendency toward linearity was suggested by plotting the mean variation within deciles for Reading Comprehension. (Author/BS)

\*\*\*\*\*  
 \* Reproductions supplied by EDRS are the best that can be made \*  
 \* from the original document. \*  
 \*\*\*\*\*

ED253564

Longitudinal Variability of Status Scores Among Students with  
Varying Levels of Achievement

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

R.L. Egelston

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)."

Richard L. Egelston  
NYS Education Department

and

Rosemarie Heinegg  
Scotia-Glenville (NY) Central Schools

U.S. DEPARTMENT OF EDUCATION  
NATIONAL INSTITUTE OF EDUCATION  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

☒ This document has been reproduced as  
received from the person or organization  
originating it.

☐ Minor changes have been made to improve  
reproduction quality.

• Points of view or opinions stated in this docu-  
ment do not necessarily represent official NIE  
position or policy.

Presented at the Annual Meeting of the  
Northeastern Educational Research Association

Ellenville, New York  
October, 1984

TM 810 104

## Longitudinal Variability of Status Scores Among Students with Varying Levels of Achievement

Hoover (1983) indicated that "... status scores, tell how a pupil's test performance compares with that of others in a single reference group .... Status within a single reference group is usually expressed in terms of percentile ranks. Other commonly used status scales, stanines, T-scores, and normal curve equivalents are obtained from one-to-one transformations of percentile ranks." In most school districts annual testing programs yield status scores for students which may be examined over time. If it is assumed that standardization samples are collected with equal care for each level of a test battery, and if it is assumed that subtests within each level of the battery sample the content domains in a similar fashion, it is possible to examine the stability of status scores over a period of several years.

These authors (1984) examined the reliability of NCE scores for Reading Comprehension and Total Mathematics subtests of the Iowa Test of Basic Skills from periods of time ranging from two to six years. Reliability coefficients for Reading Comprehension derived from annual spring measurements ranged from .85 to .94. Standard errors of measurement were small (4 NCEs), although dependent upon the estimate of true status, due to the non-linear transformation of percentile ranks to NCE's.

### Statement of Problem

While the above study focused on the entire range of student achievement in a particular district, the present study focuses upon the same students grouped by estimated true score. An initial attempt to examine reliability of status for the series of ten achievement levels proved fruitless when the variation between people was very much smaller than the variation of the interaction among people and scales. Thus, the investigation was changed to studying the within subject variation among single annual estimates of true status (as opposed to the mean of six estimates per subject). It was hypothesized that high scoring students and low scoring students would have a lower average variation among estimates (hence be more stable) than students scoring in the middle of the achievement range. Such a curvilinear relation would be required if the extreme two or three NCE values were used, just due to truncation of scale effects. But curvilinearity is not so obvious if intermediate levels of achievement are used.

### Sample

Students in a suburban school district in a northeastern SMSA who were in grade 3, and who continued through grade 8 in the same class, formed three cohort groups for this investigation. All students in each of three successive classes were administered the Iowa Test of Basic Skills (ITBS), during May, for six consecutive years. Raw scores were converted to percentile ranks, using national norms, and were then transformed to normal curve equivalents for the Reading Comprehension subtest and the Total Mathematics score. Form 6 was used within the district from 1976 through 1978, and Form 8 was used from 1979 through 1983. Appropriate grade levels of the ITBS battery were administered to each student. The three cohort groups contained 153, 148 and 148 students respectively. Due to transfers and retention, these figures represent approximately 55 percent of the total number of students in each cohort group.

Means and standard deviations were comparable for each cohort group in each subtest, and the cohort groups were combined for  $N = 449$ . When compared with the standardization sample on norms for schools, the mean performance by grade, and cohort group was in the top 20% for Reading Comprehension and in the top 10% for Total Mathematics. In fact, very few students scores in the lower two deciles of NCE performance.

### Method

The mean and variance of the six estimates of true status were computed for each student on both subtests. Pearson product moment correlations of the mean and variance yielded values of  $-.16$  and  $-.28$  for Reading and Mathematics respectively. These low, but significant correlations, suggested the possibility of a nonlinear relationship.

Mean NCE scores were then grouped in deciles of achievement according to the normal curve (e.g., any NCE mean equal to or above 76.99 was classified in the tenth decile; any NCE mean less than 76.99, but greater than or equal to 67.72 was classified in the ninth decile; etc.). For each of the ten achievement levels of each variable, the mean, median, and standard deviation of the variance within subjects was computed. Low standard deviations would show consistent variability among students. The measures of central tendency were expected to confirm or deny the hypothesis of curvilinearity.

### Results and Discussion

For Reading Comprehension there were only two students in the first decile and only six in the second decile, making any conclusions about the achievement levels very tenuous. Similarly with Total Mathematics too few cases are contained in the lower two deciles.

For Reading Comprehension the trend from the third decile through the tenth decile is generally one of decreasing average variability among students (see Table 1). The standard deviations have a slight tendency to decrease with an increase in achievement levels, although the standard deviations are similar in magnitude to the mean. Taking these descriptive statistics into account, it appears that students who perform at higher reading levels have a tendency to be more stable than do students of lower achievement levels in reading. A linear trend is somewhat apparent with this grouping.

For Total Mathematics the case is not clear cut at all. A near perfect rank order correlation exists between the mean and standard deviation of the within subject variances, and neither appear to be related to the achievement decile (although students in the tenth decile are decidedly more stable than the others).

### Conclusions

Overall there is slight, if any, tendency for higher achieving students to be more consistent in their annual status position than lower achieving students. The possible exception to this is in the extreme top decile.

It is possible to derive the maximum variance among a set of six scores if the mean score is specified. If this is done, then the maximum possible variance increases the closer the mean gets to an NCE of 50. This would suggest a quadratic relation

between achievement level and variance, but such a relation was not found when achievement levels were blocked into ten groups. Thus, the correlation of means and variance cited earlier appear to indicate a low tendency toward a linear relation, rather than a tendency toward a quadratic relation.

#### Reference Notes

Egelston R. and Heinegg, R. A comparison of the stability of percentile ranks and normal curve equivalents. Presented at the annual meeting of the New England Educational Research Association, Rockport, ME, April 1984.

Hieronymus, A. N., Linquist, E. F., and Hoover, H. D. Manual for School Administrators Iowa Tests of Basic Skills Forms 6 & 8. Boston: Houghton Mifflin Co., 1971.

Hoover, H. D. "The most appropriate scores for measuring educational development in the elementary schools: GE's." Paper presented at the annual meeting of the American Educational Research Association, Montreal, 1983.

Table 1

Descriptive Statistics for Within Subject Variance by  
Achievement Decile and Subtest Score

|                | Achievement Decile |     |     |     |     |    |    |     |    |     |
|----------------|--------------------|-----|-----|-----|-----|----|----|-----|----|-----|
|                | 1                  | 2   | 3   | 4   | 5   | 6  | 7  | 8   | 9  | 10  |
| <u>Reading</u> |                    |     |     |     |     |    |    |     |    |     |
| N              | 2                  | 6   | 20  | 22  | 46  | 61 | 77 | 73  | 77 | 65  |
| Median         | 145                | 139 | 86  | 68  | 59  | 43 | 56 | 49  | 43 | 50  |
| Mean           | 197                | 155 | 104 | 84  | 84  | 56 | 80 | 78  | 67 | 60  |
| St. dev.       | 75                 | 53  | 75  | 74  | 100 | 44 | 68 | 90  | 57 | 40  |
| <u>Math</u>    |                    |     |     |     |     |    |    |     |    |     |
| N              | 0                  | 3   | 14  | 21  | 31  | 41 | 63 | 72  | 82 | 122 |
| Median         | --                 | 189 | 50  | 106 | 82  | 69 | 66 | 82  | 72 | 38  |
| Mean           | --                 | 189 | 81  | 160 | 126 | 87 | 95 | 124 | 93 | 43  |
| St. dev.       | --                 | 18  | 62  | 192 | 98  | 70 | 97 | 147 | 78 | 30  |

## ABSTRACT

### Longitudinal/Variability of Status Scores Among Students with Varying Levels of Achievement

Richard L. Egelston - New York State Education Department

Rosemarie Heinegg - Scotia-Glenville (NY) Public Schools

The stability of annual NCE scores over a period of six years has been established in a previous study of status (position within a reference group). This study investigated the relationship between status stability and achievement level of students. Correlations between means and variances suggested a non-linear relation. When achievement levels were blocked into deciles, the within decile variation was very large when compared to the means, and obscured any relation between achievement level and within subject variation, although a tendency toward linearity was suggested by plotting the mean variation within deciles for Reading Comprehension.